

Published as a separate and in *Genetic Psychology Monographs*, 1941, 24,  
311-345.

MEASUREMENT OF THE SIZE OF GENERAL ENGLISH  
VOCABULARY THROUGH THE ELEMENTARY  
GRADES AND HIGH SCHOOL\*

---

MARY KATHERINE SMITH<sup>1</sup>

---

*New Concord, Ohio*

---

|                                      |     |
|--------------------------------------|-----|
| I. Introduction . . . . .            | 313 |
| II. Major variables . . . . .        | 315 |
| III. Purpose and procedure . . . . . | 319 |
| IV. Results and discussion . . . . . | 327 |
| V. Summary . . . . .                 | 343 |
| References . . . . .                 | 345 |

---

\*Recommended for publication by Robert H. Seashore, and received in the Editorial Office on June 28, 1941. Copyright by The Journal Press.

<sup>1</sup>The writer is deeply indebted to Dr. Robert H. Seashore for his generous assistance and constructive criticism in every phase of this study. Grateful acknowledgment is also made to the superintendents, principals, and teachers of the different schools who accorded this work their hearty coöperation.

## I. INTRODUCTION

Measurement of vocabulary has long interested psychologists because in addition to its own importance it is closely related to other psychological abilities. The rôle of vocabulary measurements has been particularly important in the measurement of general intellectual ability. Terman (7) found that the vocabulary list of the Stanford Binet correlated more highly with the results of the total test than did any three other items combined. Moreover, a list of the vocabulary tests available would be to a large extent a list of intelligence tests of which the vocabulary test is a part. We have also assumed that vocabulary bears a close relationship to the reading skills and Seashore, Stockford, and Swartz in a study on college students (5) have shown that although size of vocabulary shows no correlation with speed of reading, the number of words a person knows does correlate highly with reading comprehension.

Measurement of the size of vocabulary is not new; many estimates for different age levels have been made. However, there has been wide disagreement among these studies. The work of previous investigators has been summarized by Seashore and Eckerson (4) in a table of the size of the vocabulary from first grade through college and adult years. Some work has also been done on the pre-school level. Smith (6) constructed a vocabulary test from a sampling of Thorndike's list of 10,000 most frequently used words and from recorded vocabularies of children. She tested children between the ages of eight months and six years, using pictures and questions to elicit the use of the word itself by the child or a definition of the word. She found a range in vocabulary size from no words for the eight months old child to 2,562 words in the vocabulary of the average six-year-old, the children showing an average gain of 572 words a year.

McCarthy (3) recorded a sampling of 50 consecutive responses for children from 18 to 54 months of age. She found the total number of words used at 18 months was 20, which number increased to 230 at 54 months of age.

Smith's work is to be criticized upon the basis of the inadequacy of her testing sample, which by the nature of the test imposes an upper limit to the possible scores. Also, the criteria of knowledge which she used were too inflexible to measure the child's knowledge of the words presented.

McCarthy was interested chiefly in the nature of the words and sentences used by the child rather than the extent of vocabulary and actually measured only a partial use vocabulary. Thus her results are *not* to be taken as a measure of the complete extent of the child's vocabulary.

Among the principal factors which have brought about major variations in the estimated size of vocabulary are, first, the continual growth of the English language; second, the definition of the unit of measurement, a word; third, the criteria of knowledge employed; fourth, and by far the most important, the basis for sampling, e.g., the size of the dictionary or the nature of the use situation from which the sampling for the test has been taken. Consequently, estimates of the size of vocabulary made on the basis of counts of words employed in some writer's works or counts of words used in conversations are inadequate as estimates of total vocabulary. Likewise a vocabulary test which is based on too small a sampling of words in the English language is inadequate. A test which is based on a pocket-dictionary automatically limits its sampling to only a portion of the total number of words available. Seashore and Eckerson's table (4) shows clearly that the larger the dictionary used as a basis for a sampling, the larger the estimated vocabulary will be. Hence we can surmise that past estimates of the size of vocabulary, based on dictionaries smaller than the unabridged volumes now in use, have been too small. Seashore and Eckerson (4), in fact, have shown that the number of words which the average university student knows is much greater than previous studies have indicated, averaging about 156,000 words for undergraduate students.

In the light of the variety of methods used and the results obtained it is well to consider the major variables of the measurement of vocabulary.

## II. MAJOR VARIABLES

### A. DEFINITION OF A WORD AS A UNIT OF MEASUREMENT

First, we need to define the unit with which we work, the word. We shall define a word as, "a dictionary item," as Seashore and Eckerson (4) have defined it in the construction of their vocabulary test. They make the distinction between "basic" and "derived" terms. Basic words are those which in the dictionary are printed in heavy type as separate entries along the margin. Derived terms are compound terms or words formed from the basic word, "usually listed in medium type and indented under the basic term" (4). For example, "loyal" is a basic word, "loyalize" and "Loyal Legion" are derived terms. Neither additional meanings for a word nor variant spellings are counted as separate words, but the same stem used in different parts of speech, when listed separately, and compound terms are counted as separate words under such a definition. The dictionary seems to follow the practice of listing separately different forms of the same word when *in that part of speech* it has a meaning not clearly indicated by the ending added to the word, e.g., "fixing" in the meaning of a dye chemical is listed separately from the verb "fix."

On the other hand, Thorndike (8) would classify as variations of the main word all words which are:

- a) plurals, *s* or changing *y* to *ies*
- b) Adverbs, *ly*
- c) comparatives and superlatives, *er* and *est*
- d) verb forms, *s*, *d*, *ed*, *ing*
- e) past participles adding *n*
- f) adjectives, adding *n* to proper nouns

This definition of Thorndike's gives fewer "words" in the language than Seashore and Eckerson's definition, but many more, perhaps 50, types of variations could also have been included which would decrease the number of separate words still more.

Table 1 shows a comparison of the number of words on sample pages of Funk and Wagnalls *New Standard Dictionary* (unabridged) as counted according to Thorndike's definition and as classified by Seashore and Eckerson (4). The average number of "words" as counted according to Thorndike's definition is 77 while the number of words as counted according to Seashore and Eckerson's definition of a basic word is 60. Thus, a statement of the

TABLE 1  
 COMPARISON OF THE NUMBER OF WORDS IN FUNK AND WAGNALLS *New Standard Dictionary* ACCORDING TO DEFINITIONS BY THORNDIKE AND BY SEASHORE AND ECKERSON

| Vol. 1<br>Page<br>number      | Thorndike<br>Units<br>Number<br>of words | Seashore and Eckerson Units                   |   |                   |
|-------------------------------|--|---|---|-------------------|
|                               |  | "Basic"<br>Marginal<br>words in<br>heavy type | "Derivatives"<br>Other parts<br>of speech | Compound<br>words |
| 101                           | 77                                       | 70  | 17  | 33                |
| 201                           | 39                                       | 34  | 11  | 112               |
| 301                           | 52                                       | 58  | 6   | 95                |
| 401                           | 86                                       | 40  | 27  | 91                |
| 501                           | 59                                       | 62  | 11  | 36                |
| 601                           | 45                                       | 55  | 9   | 97                |
| 701                           | 44                                       | 38  | 16  | 32                |
| 801                           | 88                                       | 40  | 22  | 50                |
| 901                           | 54                                       | 49  | 19  | 67                |
| 1001                          | 134                                      | 89  | 45  | 19                |
| 1101                          | 152                                      | 107   | 15  | 28                |
| 1201                          | 116                                      | 53  | 39  | 27                |
| 1301                          | 146                                      | 47  | 43  | 58                |
| 1401                          | 49                                       | 67  | 6   | 16                |
| Totals                        | 1141                                     | 809   | 286                                       | 751               |
| Ave. Vol. I                   | 81.4                                     | 57.7  | 20.4                                      | 53.6              |
| Ave. Vol. II                  | 73.5                                     | 62.9  |   |                   |
| Ave. I + II                   | 77.4                                     | 60.3  |   |                   |
| x 2757 pages<br>in dictionary | 213,392                                  | 166,247                                       | 56,243                                    | 147,775           |

size of vocabulary in terms of basic words appears to be a conservative estimate. On the basis of this sampling of the dictionary it would seem that vocabulary in terms of Thorndike's units is about 25 per cent larger than the basic vocabulary measured by Seashore and Eckerson, and about 42 per cent smaller than their total vocabulary. In other words, basic vocabulary scores from the Seashore-Eckerson test may be converted approximately into Thorndikean units by multiplying them by 1.28.

#### B. CRITERIA OF KNOWLEDGE

Criteria of knowledge which may be applied to vocabulary are: (a) recognition of the commonest meaning of a word, (b) definition in the subject's own words, (c) use of the word in a sentence or citing an illustration.

### C. RELATIVE VS. ABSOLUTE SIZE OF VOCABULARY

We may make a distinction between size of vocabulary as reported in absolute and in relative terms. By "absolute" we mean the total number of words in one's vocabulary as estimated from a representative sampling of all the words in a dictionary. By "relative" we indicate the size of one's vocabulary in relation to the vocabularies of some group of persons. The absolute vocabulary may not be known for a single person in the standard group, but the size of vocabularies are known in relation to each other. In indicating the size of a person's vocabulary by this second method we often make use of such approximations as quartiles, deciles, and centiles which have been derived from the standard group.

It is possible to measure size of vocabulary in terms of how many words an individual knows on an arbitrarily selected list in comparison with the number of such words that other individuals know. Gansl (1) preferred to use such a relative score rather than to try to measure the absolute size of vocabulary in thousands of words, since so many different results had been obtained in absolute measurements prior to the publication of Seashore and Eckerson's analysis of the factors involved in such measurements. She was interested in the growth of vocabulary through elementary school and a relative score served her purposes fairly well, although short samplings from limited sources may impose an artificial upper limit in scores.

Secondly, we may get an absolute score, an estimate of the actual number of words in the vocabulary. For this purpose the test used must not only be reliable, but must be made from a large enough sampling of the words in the English language to give the best person tested a chance to show all he knows.

There is need for having some idea of absolute size of vocabulary at the elementary and high school levels. Knowing the number of words which an individual can use correctly gives some idea of the breadth of his information and of his intellectual tools.

Since children's books for the lower grades especially are being built around carefully controlled reading vocabularies and introduce only a limited number of words per book, it will be significant to know how many words the beginning pupil may be expected to know and how many words he normally will add to his store in a year's time. To be sure, this limiting of the number of words introduced to the child as he is beginning to learn to read is a good teaching

device from the point of view of mastering a technique. There is a great discrepancy at the first grade level between the number of words which the child *can read* and the total number he can *recognize when spoken*. There probably are several different vocabularies such as reading and pronouncing, or reading and comprehending, as well as general comprehension vocabulary. In later grades, even when reading vocabulary has increased spelling vocabulary is still rather limited.

It is a problem for future research as to when and how these discrepancies disappear, for Seashore and Eckerson (4) have shown that at the university level there is little difference in the size of vocabularies as measured by any of the three criteria of knowledge. Individuals were able to use and illustrate about 92 per cent of all the words they could recognize.

We are interested not only in the nature of the growth of the size of vocabulary through the school years, but also in an estimation of the actual number of words which children of those ages can use. An absolute measurement in terms of the total number of words known can also be translated into relative terms such as centiles, deciles, and quartiles.

The test used in this study was the *English Recognition Vocabulary Test*<sup>1</sup> by Seashore and Eckerson (4). It has several advantages for our purpose of securing absolute scores. First, it was constructed from a sampling of Funk and Wagnalls' *New Standard Dictionary of the English Language*, two volume edition of 1937, one of the largest available at that time. Second, the items in the main part of the test are presented in a multiple-choice form which meets the criterion of knowledge as recognition of the commonest meaning of the words. Third, words in the test are arranged in order of difficulty which allows for testing on only the first part of the test with the younger children. The scoring formula of the test permits an estimate of vocabulary size from whatever number of words an individual has attempted to define.

---

<sup>1</sup>Published by the authors, Evanston, Illinois.

### III. PURPOSE AND PROCEDURE

#### A. PURPOSE

1. To determine the criteria of knowledge and the procedures necessary for measuring the absolute size of vocabulary among subjects below college level and particularly at the lower ages.
2. To employ these procedures in determining the individual differences in size of vocabulary over the range from first to twelfth grades inclusive.
3. To determine the central tendencies and variabilities of these measurements for a description of the growth of vocabulary during this period.
4. To provide tentative norms in the absolute size of vocabulary for this range of age and grade levels.

#### B. PROCEDURE

The *English Recognition Vocabulary Test* by Seashore and Eckerson (4) was given to pupils from first grade through high school in two schools<sup>2</sup> and through the first eight grades in a third school.<sup>3</sup>

There are three parts to this test. The first part consists of 173 multiple choice items made up of basic general terms, arranged in approximate order of difficulty. The second part contains 158 words which are either proper nouns or rare words. These words are compactly printed in four columns and from this number the student is expected to choose those few words which he knows and to write definitions for them. The third part has 46 "derived terms," that is, variations in parts of speech as well as compound and technical terms, also arranged in order of difficulty. For part three as well as in part two the subject must write out the meaning of the words. Since relatively few words are ordinarily known in parts two and three of the test, this written portion of the test is rather brief.

For adults the test may be used as either a speed test or a power test. For work with children and especially in investigating the absolute size of vocabulary, it is necessary to use the test without time limits, as a power test.

---

<sup>2</sup>New Concord-Union Rural School, New Concord, Ohio, tested Dec. 10, 1939, to Jan. 5, 1940, and Northbrook Public Schools, Northbrook, Illinois, tested Feb. 5, 1940, to March 5, 1940.

<sup>3</sup>Cleveland Public School, Niles Center, Illinois, tested March 6, 1940, to April 3, 1940.



In order to measure absolute size of vocabulary the method of testing must be adapted to the purpose at hand which is to ascertain *for how many of the words in the test-sampling the child knows some correct meaning*. It would be relatively easy to adopt one criterion of knowledge and to measure the children's performance upon the vocabulary test in accordance with that criterion and thus secure rankings in vocabulary ability. However, in measuring absolute size of vocabulary we do not want mere rankings; we desire an adaptable measuring method which will allow the child to show what he knows about the words in the test. This means the eliminating as far as possible of the influence of such factors as the child's ability to read and to spell, because we are interested only in measuring the number of words for which the child has some effective knowledge and not in his expression of that meaning.

Preliminary testing indicated that a combination of several criteria of knowledge would best accomplish this purpose, especially in testing at the lower elementary levels. In fact, there is an increasing stringency of the criteria of knowledge of a word with increase in the chronological age of the subject. This difference in criteria at the two age levels cannot be equated or legitimately ruled out except through individual testing throughout the grade range, because children simply do improve in both the quality and the quantity of words known. It was our purpose to discover those methods of testing at the different grade levels which would give the most accurate picture of the growth of vocabulary.

Preliminary tryouts showed that there were three main administrative groups into which the subjects fell: (a) Early elementary, grades one, two, and three; requiring individual, oral testing. (b) Middle elementary, grades four, five, and six; requiring aid in reading. (c) Late elementary and high school; requiring only opening directions and occasional supervision and aid in reading. However, the change from one method to another in testing different grade levels is gradual and adaptations for exceptional individuals must be made at all levels.

In late elementary and high schools the test was administered to groups of from 25 to 40 students. The examiner always introduced the work by telling the students that the test was part of a study which was for the purpose of finding out how many words high school and grade school people knew. They were assured that the

results of the test would have no bearing on their school grades. The high school testing was done mainly through English classes. They were told the test was one which could be used from first grade through college; that, accordingly, it began with easy words and gradually became harder. They were cautioned against carelessness on the easy words and encouraged to guess on any words which seemed at all familiar.

After the students had filled in the data on the front of the test, they were asked to open the booklets; were shown the example at the top of the page; and permitted to start work. The examiner encouraged them to attempt all words they knew, by saying, "*It is a good idea to guess. If you don't know a word at all, leave it out, but if you can make a good guess, be sure to try the word.*" The examiner also offered to define any of the words used in the multiple choice responses and to pronounce any of the words. The students were told, "*If you don't know what one of these words in light type means, ask and I will tell you. If you want any word pronounced, ask and I will pronounce any of the words for you.*"

The examiner then made it a point to move among the seats and to make sure that each student was following directions as to the placing of the numbers. This also afforded the students an opportunity to ask any questions they might have.

When the first several pupils had completed part one, the examiner asked for the attention of all of the group. The words on part three were pointed out, their nature explained, and directions for writing out the definitions given. The examiner said, "*The words in part three are compound and technical words. You are to write out what they mean. Be sure to write a full explanation or description; tell enough so someone else could tell just what you mean. For example, number one is 'for mercy's sake'. Just explain what you mean when you say, 'for mercy's sake'. Write the meanings for as many of the words as you can.*"

A little later the group as a whole was shown part two and given these instructions: "*Part two has words which are rare or which are the names of places and people. You are to write out what the words mean for as many words as you can. Since there are so many words you might notice particularly these words. Check the ones you know and then write out the meanings when you are ready.*" (Here the examiner read the number and pronounced the word for 27 of

the words.) "You may write out any other words you know, too. Finish part one and part three, and then do part two."

The test was administered in classes which had 50 to 45 minute periods. Those students who did not finish in one period were given an opportunity to complete their papers at the next meeting of the class.

The first part of the test was scored on the basis of the number of words which the pupil attempted to define. From the number attempted was subtracted the errors corrected for guessing and the resulting score was multiplied by a constant which gave the number of basic words known. Scores on parts two and three were the number of words correctly defined multiplied by constants. In scoring parts two and three credit was given only for definitions which indicated the meaning of a compound term *as a whole* and not definitions of the separate parts. Half-credits were used liberally. The sum of these three scores represented the total vocabulary.<sup>4</sup>

In the middle elementary grades, four, five, and six, the children were given the same introductory remarks as those used for the high school. In addition the examiner said, ". . . *We will work together. I will read the words for you. Now the first word in dark type is 'adhesive.' If adhesive is 'slippery' put a 1 in front of it. If adhesive is 'rough', put a 2 in front of adhesive. If adhesive is 'fatty' put a 3 in the parenthesis. Or if adhesive is 'sticky' put a 4 in front of adhesive. Now don't say anything about the words; just mark it.*" After making sure that each child was marking this word correctly, the examiner continued. "*The second word is 'quick.' Does quick mean the same thing as 'dead'? Does quick mean 'fast'? Does quick mean 'good'? Or does quick mean 'slow'? Put the number of the word that means the same thing as quick in front of quick.*" Then with less elaborate questions, the examiner continued to read each test word, repeating the word to be sure each pupil heard correctly and then read, in order, the four possible answers, avoiding cues from the voice. Whenever the children requested them, definitions for the choice words were given.

Standard definitions had been written out before the work of testing began, so that the examiner furnished the same answers whenever requested for a definition. On certain items, the examiner

<sup>4</sup>More detailed scoring directions are printed on the test blank and are discussed by Seashore and Eckerson (4).

did not wait for the children to request definitions, but incorporated them into her original question, as: "*Number 14, takedown. What do we mean by takedown? Does takedown mean reputation, that is what people say about you? Or does takedown mean lowering? Is takedown, celebration? Is takedown, honors?*"

The range in difficulty of the words in this test is so great that the last page contains words which are almost all too difficult for pupils of the fourth, fifth, and sixth grades. In addition, since the alternative responses for the more difficult items were so chosen as to test a college student's ability to distinguish the correct meaning from those which were fallaciously similar in sound, spelling, etc., younger students may obtain somewhat *less than chance* scores on this portion of the test when items on the last pages are marked blindly. This is an instance of a further qualitative difference in the nature of vocabulary knowledge at different ages.

Consequently it is wise to suggest to the children that they leave out any words on pages three, four, and five which they have never seen nor heard before. In general, fourth and fifth graders should attempt at least 100 words; sixth, seventh, and eighth graders at least 110 words. The specific words which the examiner should urge the children to attempt are listed in the manual for upper elementary grades.<sup>5</sup>

There will be children who will mark all the items in the test regardless of instructions. Any child who wants to should be allowed to mark the items in order to avoid penalizing the best students. However, in the scoring only the designated 100 or 110 words should be considered unless it is found that the score based on the entire test is higher; in this case the higher score is taken as the one representative of the child's knowledge.

For parts two and three the children in grades five and six were requested to write the meanings of the words as in the upper grades.

In grade four additional precautions were taken. Only four to six children were tested at a time on part one, in the manner described. And for parts two and three, each child was taken individually. The examiner gave the child an unmarked test booklet and then asked the child to define the words on parts three and two. The examiner wrote the child's reply as he gave it, encouraged

---

<sup>5</sup>This manual may be secured from the author.

him to attempt all the items he could, and asked whatever questions were necessary to clear up ambiguous replies.

In addition, the children were told that some of the words in part one had two meanings and that for some of them the examiner wanted examples. The examiner then asked the child to give examples, to use in sentences, and to define in his own words, and in some cases to give additional meanings for those words which the child had marked incorrectly, but which most third grade pupils could define. Similar follow-up questioning was used with certain fifth and sixth graders who had missed words which children in the first three grades answered correctly.

In grades one, two, and three the children were tested individually. The examiner asked the child to define the word. If the child could not reply at once or if he did not make his meaning clear, the examiner read the choice words. Preliminary try-outs showed that these four alternative responses if read in a series made too great an amount of material for the child's memory span. The younger child could not comprehend such a complicated question. For this reason the alternative responses were phrased in separate short questions which repeated the test word with each alternative response. These could be read one right after the other without exceeding the child's memory span. The questions which were used by the examiner were prepared in a manual and kept standard for each child.<sup>6</sup>

In some instances the child's initial response to the word alone was incorrect, but when the alternative responses were presented he corrected his first response and chose the correct answer. In this case he was given full credit for the word, allowing of course for chance.

The child sometimes gave stereotyped answers to the alternative responses on the basis of position cues; that is, he would always choose the last of the four alternative responses or the first, without regard to their meaning. In such a case or if the child did not respond to the multiple-choice question at all, he was asked to describe the object in his own way or to tell something about it, or to use the word in a sentence and then to explain his sentence. If his reply to such questions was correct the choice of an incorrect alternative response was disregarded and he was given credit for that word.

---

<sup>6</sup>This manual may be secured from the author.

When all other questions failed, leading questions were asked and only half-credit was given for correct replies to such questions.

For all of the words, any correct meaning which the child could give was credited. For some words the child might choose the wrong alternative responses or say that he did not know the word and yet be able to define the word correctly in terms of another meaning, in which case he was given full credit. For example, the child might not know "poker" as a game, but know it as a fire-tool.

#### IV. RESULTS AND DISCUSSION

Since there are two common conceptions of a "word," the scoring of the vocabulary test is arranged to allow for analyzing the results into terms of either basic words alone or of total vocabulary which includes both basic and derived terms. The results then may be quoted in those terms which fit the reader's preference in the definition of a word.

Before examining quantitative differences it is well to remember that the qualitative criteria of knowledge becomes increasingly stringent with progress through the grades and that to this extent the scores are not strictly comparable. However, one would scarcely expect the quality of meanings to be the same at widely different ages and no statistical allowances would seem to be called for. It is simply an observed fact that vocabulary grows both qualitatively and quantitatively.

Figures 1 and 2 show the means, quartiles, and range of the basic and total vocabulary scores for grades one to twelve inclusive.

The number of pupils tested were: 44 in first grade, 40 in second,

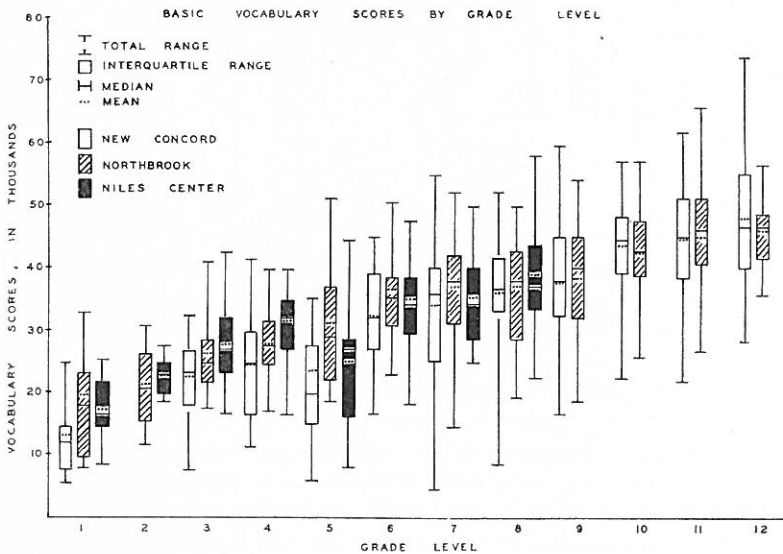


FIGURE 1

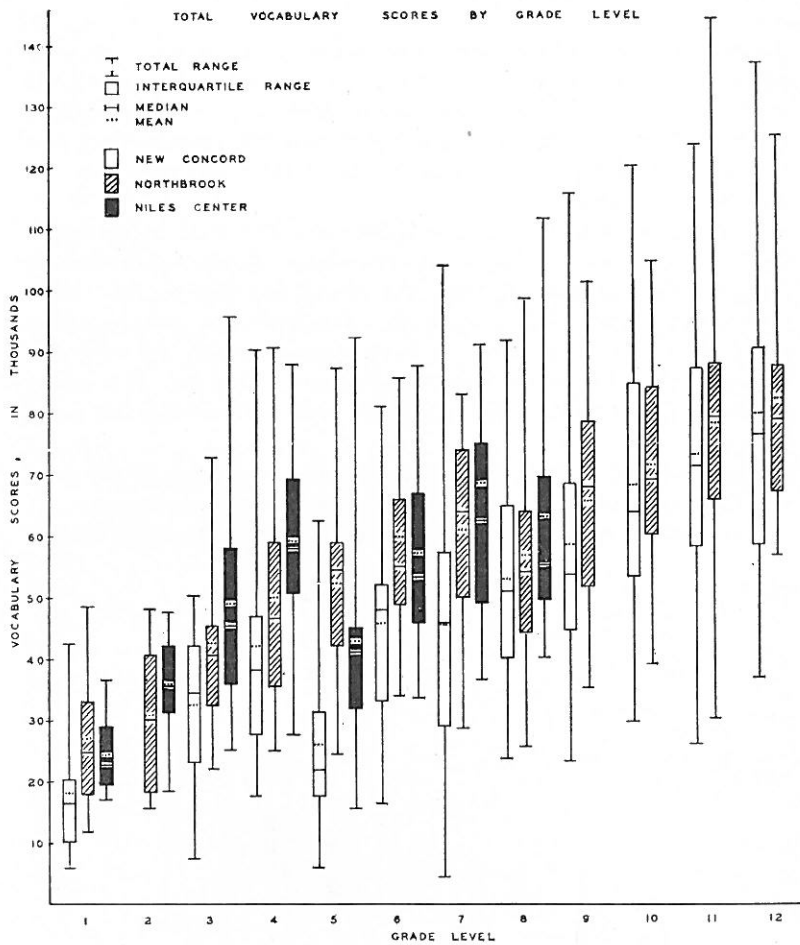


FIGURE 2

59 in third, 73 in fourth, 61 in fifth, 66 in sixth, 69 in seventh, 71 in eighth, 114 in ninth, 111 in tenth, 95 in eleventh, and 64 in twelfth grade; making a total of 867 pupils tested.

It will be noted that the average size of vocabulary whether analyzed for basic words alone or by a total score including derived words is numbered in the tens of thousands. When we give the



child an opportunity to show what he knows about an adequate sampling of words and avoid restricting his performance by inadequate testing methods or by a sampling which imposes an artificial ceiling the child gives evidence of knowing a great many more words than we have hitherto estimated that he knew. Absolute size of vocabulary throughout the grade distribution greatly exceeds past estimates.

One of the striking characteristics of these distributions is the variability in scores within one grade. The range of individual differences in raw scores, especially for the total vocabulary is very great. Table 2 gives the percentage of variance (one-half the inter-

TABLE 2  
PERCENTAGE OF VARIANCE ( $Q/M$ ) OF SCORES IN EACH GRADE FOR BASIC AND TOTAL VOCABULARY SCORES

| Grade | New Concord |       |          | Northbrook |       |          | Cleveland<br>Niles Center |       |          |
|-------|-------------|-------|----------|------------|-------|----------|---------------------------|-------|----------|
|       | Basic       | Total | <i>N</i> | Basic      | Total | <i>N</i> | Basic                     | Total | <i>N</i> |
| 1     | 26          | 27    | (13)     | 21         | 28    | (19)     | 20                        | 20    | (12)     |
| 2     |             |       |          | 26         | 36    | (18)     | 11                        | 15    | (20)     |
| 3     | 20          | 29    | (17)     | 13         | 15    | (25)     | 16                        | 23    | (16)     |
| 4     | 27          | 23    | (18)     | 12         | 24    | (25)     | 12                        | 16    | (30)     |
| 5     | 29          | 26    | (21)     | 24         | 16    | (22)     | 25                        | 15    | (18)     |
| 6     | 19          | 21    | (22)     | 11         | 15    | (25)     | 13                        | 18    | (19)     |
| 7     | 23          | 31    | (29)     | 15         | 20    | (21)     | 16                        | 19    | (19)     |
| 8     | 12          | 24    | (29)     | 19         | 17    | (22)     | 13                        | 16    | (20)     |
| 9     | 17          | 20    | (70)     | 17         | 20    | (44)     |                           |       |          |
| 10    | 10          | 38    | (64)     | 10         | 17    | (47)     |                           |       |          |
| 11    | 14          | 20    | (61)     | 11         | 14    | (34)     |                           |       |          |
| 12    | 16          | 20    | (42)     | 8          | 12    | (22)     |                           |       |          |

quartile range, or  $Q$ , divided by the mean) of the scores within each grade. With the exception of the fifth grade, the percentage of variance is greater for the total vocabulary scores than for basic vocabulary.

It will be noted, too, that there is great over-lapping of scores from one grade to another. The over-lapping of basic vocabulary scores is so great that only grades one and two are completely outside the total range of individual differences in grade twelve. This is true only if the grades are compared within any one school system. Comparing all the schools together there is over-lapping of even grades one and twelve. Within any one school system the highest first and second graders knew more basic words than did the

poorest student in every other grade level up to and including the eleventh grade! In the scores for total vocabulary the overlapping is as great. The twelfth grade's lowest score exceeds the highest scores in only first and second grades in the Northbrook School, while first and twelfth grades overlap in the New Concord School. However, in the Cleveland School the highest first grader does not equal the lowest eighth grader in total vocabulary scores although the two grades overlap in basic scores; such fluctuations are, of course, greatly influenced by a few extreme scores.

This extreme overlapping is reduced somewhat when we compare quartiles rather than total ranges. In scores on basic vocabulary,  $Q_1$  of the high school seniors exceeds  $Q_3$  of the first six grades, while  $Q_1$  of the eighth grade exceeds  $Q_3$  of the first three grades. It is much the same for total vocabulary;  $Q_1$  of the eighth grade exceeds  $Q_3$  of only the first two grades, while  $Q_1$  of the high school seniors is higher than  $Q_3$  of the seventh grade in the New Concord and of the sixth grade in the Northbrook School. *It seems we have underestimated the ability of our better students and overestimated the ability of the poorer students all through the school.*

There is a progressive growth in the average size of vocabulary from grade to grade although the rate of increase is not very regular. Gansl (1), who measured relative size of vocabulary from grade three to eight inclusive, found that "the age-progress curve for vocabulary in this age range is best described as a straight line, with a slight tendency toward negative acceleration between the ages of twelve and thirteen." This description does not fit our results very well. The nature of her sampling of test words was such that the negative acceleration at grades seven and eight may be due to an artificial ceiling in the test. Also, her test was given as a written group test at all ages, which would introduce other factors into the results in addition to vocabulary knowledge.

In Figures 1 and 2 it can be seen that in the two schools New Concord and Cleveland, the mean for the fifth grade fell below the mean for the fourth grade. This is not true of the fourth and fifth grade scores of the Northbrook School, although there too, the difference in the two means is not as great as the difference in the means of most other two adjacent grades. There may be several reasons for this discrepancy.

One obvious explanation would be a difference in the ability of the

children. At Northbrook and Cleveland Schools the two grades were given the Kuhlmann-Anderson group test and the mean *IQ* scores for both grades were slightly over 100. Likewise the Stanford Achievement Test scores from the Cleveland School show the fourth and fifth grades both to be above their grade norms, the fifth a little more so than the fourth. For these two schools at least there is apparently no difference in the general intellectual ability of the children. At New Concord no intelligence or achievement scores were available, though the fifth grade contained many more repeaters than the fourth did and it was the feeling of the principal of the school that the fourth grade was a superior group for that school.

Another explanation may be in the change of the method of administration of the test at the fifth grade level. The fourth grade was allowed to give definitions for the words in parts two and three orally, with the examiner recording their answers. From fifth grade on the children were asked to write out the meanings of the words. There are two disadvantages to having the children write out the definitions. In the first place, the children are not facile enough in spelling and general written expression to be able to write all they know or to make their definitions exact and clear. The duller child at the fifth and sixth grade levels may turn in a paper with only two sketchy attempts at definition written on it. If this same pupil is taken individually and asked to give oral explanations for the words he may succeed in defining five or six of the terms; he is certain to do better than his written work indicated.

In the second place, it requires much encouragement from the examiner to persuade some of the children to attempt more than one or two of the words. Some of the children were so greatly impressed by the number of unfamiliar and difficult words that they gave up on the lists as a whole and failed to attempt the ones they did know. Particularly if the student was growing at all tired toward the end of the test, it was easy for him to overlook words for which he could give at least a partial meaning when questioned specifically on those words.

For this reason we recommend that the fifth and sixth grades be treated like the fourth and be tested individually and orally on parts two and three of the test.

Figures 3 and 4 present the medians, quartiles, and ranges of scores for basic and total vocabulary at year intervals of chrono-

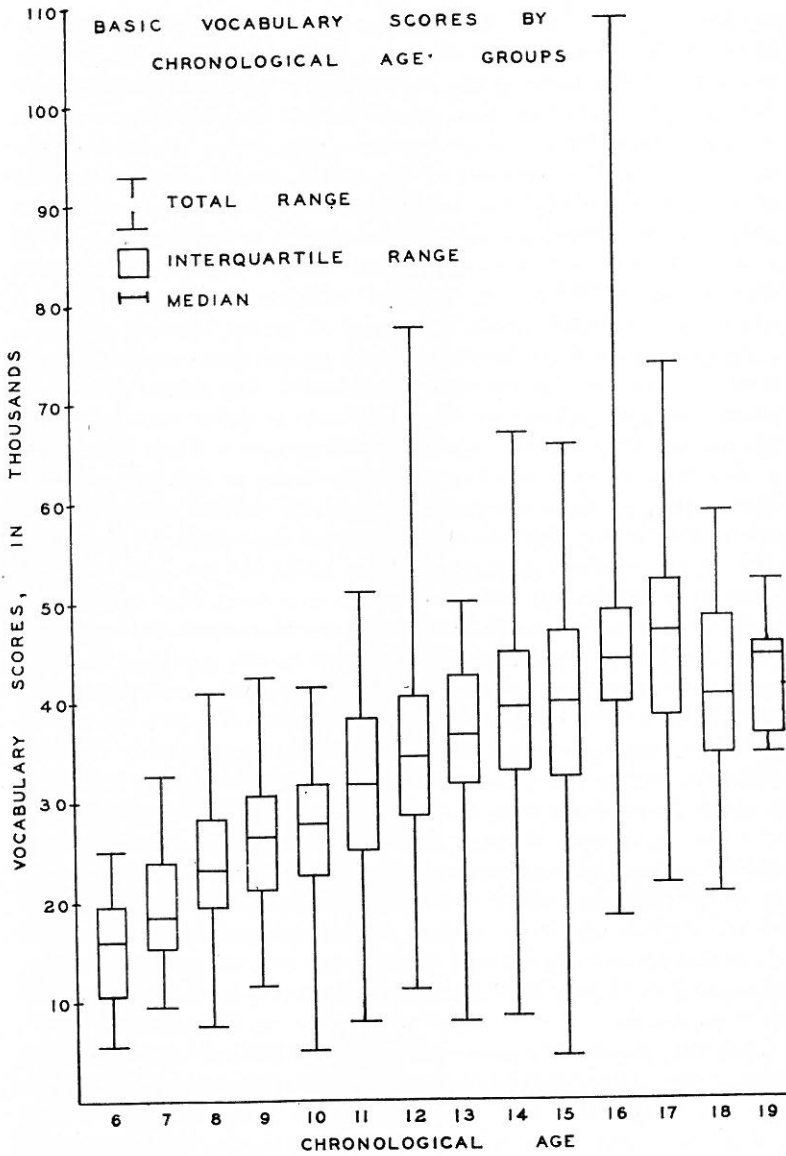


FIGURE 3

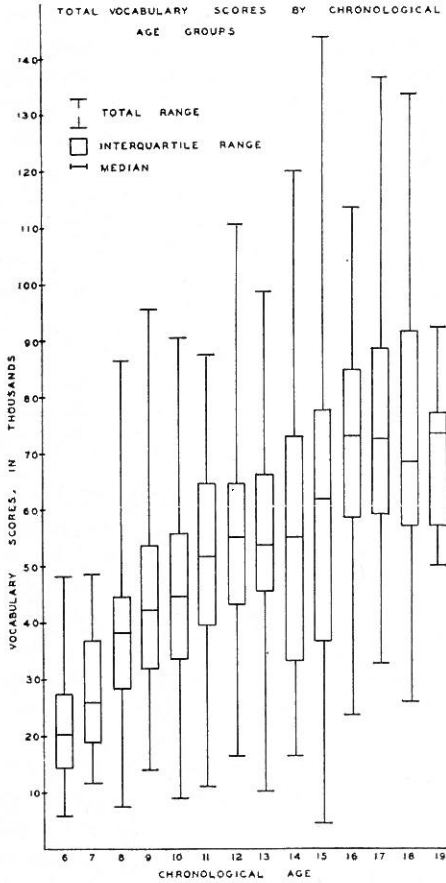


FIGURE 4

logical age which was defined in terms of the nearest birthday so that any given year represents the midpoint of a distribution, e.g., a chronological age of six means ages from five years seven months to six years six months, etc. It must be noted that the sampling for these chronological age groups represents the school population and not chronological age groups at large. This caution is particularly applicable to the upper and lower limits of the distribution. The drop in the chronological age curve at the 18-year level is probably to be explained on the basis of this sampling. Those people who are 18

years old and over and who are still in high school normally may be slightly retarded and could be expected to represent the lower end of the distribution of scores at that chronological age level. The apparent rise in the curve again at the 19-year level is a chance fluctuation in that in this sampling there were only nine subjects who were that old.

TABLE 3  
PERCENTAGE OF DERIVED WORDS IN TOTAL VOCABULARY AT THE DIFFERENT  
GRADE LEVELS

| Sch. grade | Per cent | Sch. grade | Per cent |
|------------|----------|------------|----------|
| 1          | 29       | 7          | 37       |
| 2          | 33       | 8          | 35       |
| 3          | 36       | 9          | 37       |
| 4          | 44       | 10         | 38       |
| 5          | 34       | 11         | 40       |
| 6          | 36       | 12         | 41       |

Table 3 presents the proportion of derived terms in the total vocabulary. There is a gradual increase in this proportion from first grade to twelfth. The older children not only know more words, but they are better able to handle words in general. Seashore and Eckerson's (4) study of university students showed that at that level there was a still greater difference in the number of words in the mean basic vocabulary and the mean total vocabulary. This qualitative growth in the vocabulary should be made the topic of further research.

Scores on part two, the rare words, are uniformly very low. The individual's environment and particular experience are determining factors in his knowledge of such words. It was notable that most of the older children and many of even the early elementary students in the Northbrook School knew where and what Ft. Sheridan is, as it happens to be an army training post on the lake front north of Chicago and within 10 miles of the town. Only a few of the younger pupils in the Cleveland School knew Ft. Sheridan; Niles Center is farther inland from Northbrook and is about 15 miles from Ft. Sheridan. However, at New Concord none of the children, in either the elementary or the high school, could locate Ft. Sheridan. Also an example from the derived words, the children at New Concord in the high school, many of whom come from farm homes, could explain "green manuring," which is a method of enriching the

soil by plowing under leguminous crops; whereas the Northbrook High School students did not know the term.

The results of this testing also have methodological significance. It is evident from the great overlapping of scores at the different grade levels that there is no sharp break in the general ability of the children at any point in the elementary school. In general, however, for a satisfactory measurement of absolute size of vocabulary it is best to test children in the first three grades individually, allowing the child to answer the questions orally. In the fourth and fifth grades it is well to test the children in groups of not more than four or five on part one, and individually on parts two and three. From grade six up the children may be tested in groups of 20 or 30 for part one; the sixth grade being tested orally on parts two and three, seventh and eighth writing out the definitions. Through grades six, seven, and eight the examiner should read the words for the children and may suggest which words they should attempt.

At all grade levels there will be exceptional children. Some third graders could write their answers to parts two and three very well, but on the whole that age level is not able to spell and write well enough to do their best at such a task. There are also children in fourth, fifth, sixth, and even seventh and eighth grades who should be questioned orally and individually on the words on parts two and three, and even for the words on part one in order to secure a truly representative measurement of the absolute number of words which they know. More detailed directions for the giving of the tests at the different grade levels have been worked out as manuals for the Seashore-Eckerson *English Recognition Vocabulary Test*.

It was easy to secure the cooperation of the subjects in the vocabulary testing. Children in the first four grades were eager to have their turn in taking the test. The older children were interested in the test and in their scores.

It is apparent that the three schools tested in this study, on the elementary grade level at least, are not equal. The three communities are of about equal size, ranging from a population of a little over 1,000 at New Concord and Northbrook to 5,000 at Niles Center. There were greater differences in the size of the surrounding communities and in the occupations of the parents than in the size of the school communities themselves. At New Concord there are no great cities nearer than Cleveland, Ohio, and Wheeling, West

Virginia, which are respectively 100 miles and 60 miles away. City life and even town life does not affect many of these children very much. Children at Northbrook and Niles Center, on the other hand, live on lines of direct transportation to the city of Chicago and, at Niles Center particularly, are in contact with the city frequently. The parents of the children in the New Concord School include some professional men as the town is a college community, but the greater part of the children in the early elementary grades come from farm homes. The parents of the children in the early elementary grades are also farmers, artisans, and sub-professional men to a great extent. Niles Center is more of a residential suburb for the city and a good many of the parents commute to Chicago for their work.

Thus the differences in the scores of the three schools may be due

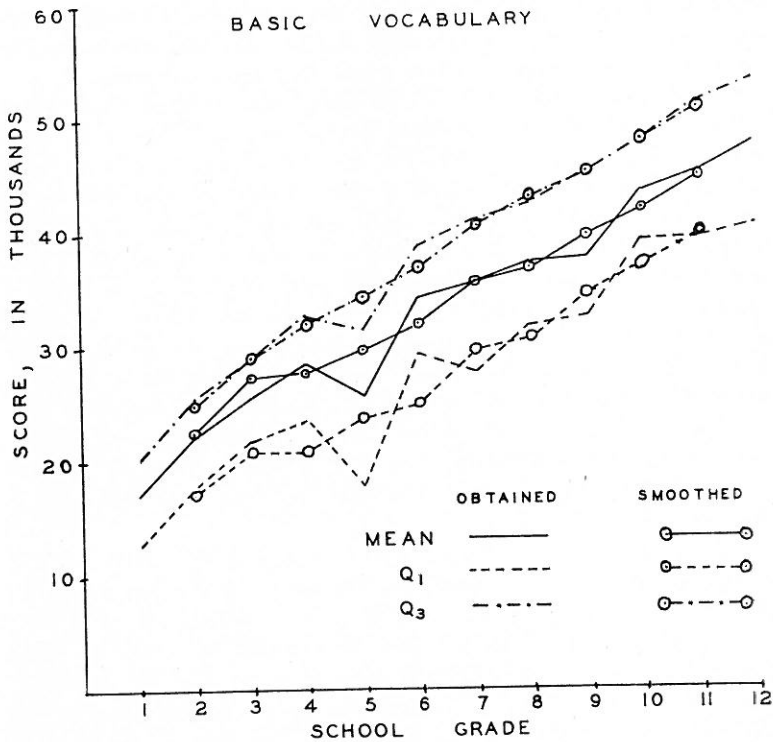


FIGURE 5



to differences in the cultural background of the pupils. It will be noted that these differences disappear at the upper grade levels. There the two schools involved sample about the same type of population, with the Northbrook High School students still having the advantage in the amount of contact with city life. It is not claimed that these three schools represent a normal sampling of American schools, but there is nothing to indicate that they are in any way atypical. Consequently it seems justifiable to average the scores at each grade level from all three schools.

The mean and quartile scores from the three schools combined

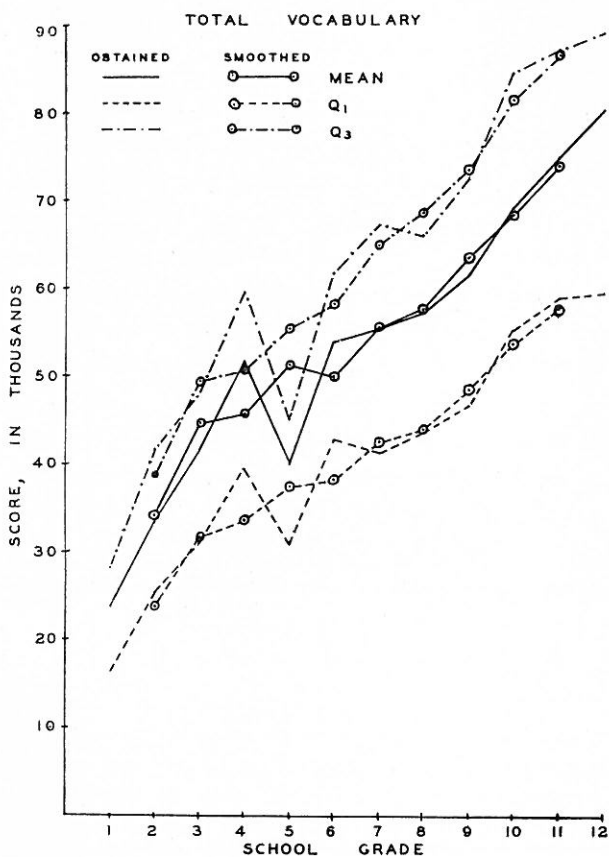


FIGURE 6

are plotted in Figures 5 and 6. Alongside these obtained means and quartiles are plotted smoothed curves, representing the score for each grade averaged with the two adjacent grades, e.g., the score at grade two on this curve represents the average of scores for grades one, two, and three. Using these two curves as a guide, another set of curves has been drawn, fitted by inspection, and shown in Figures 7 and 8. Tentative grade norms should be read from these fitted curves.

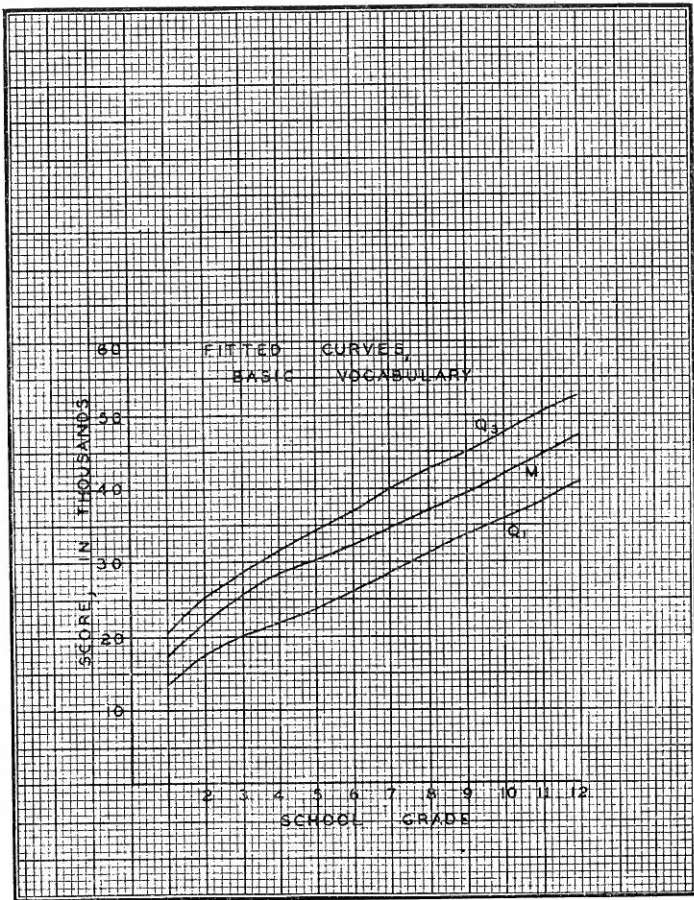


FIGURE 7

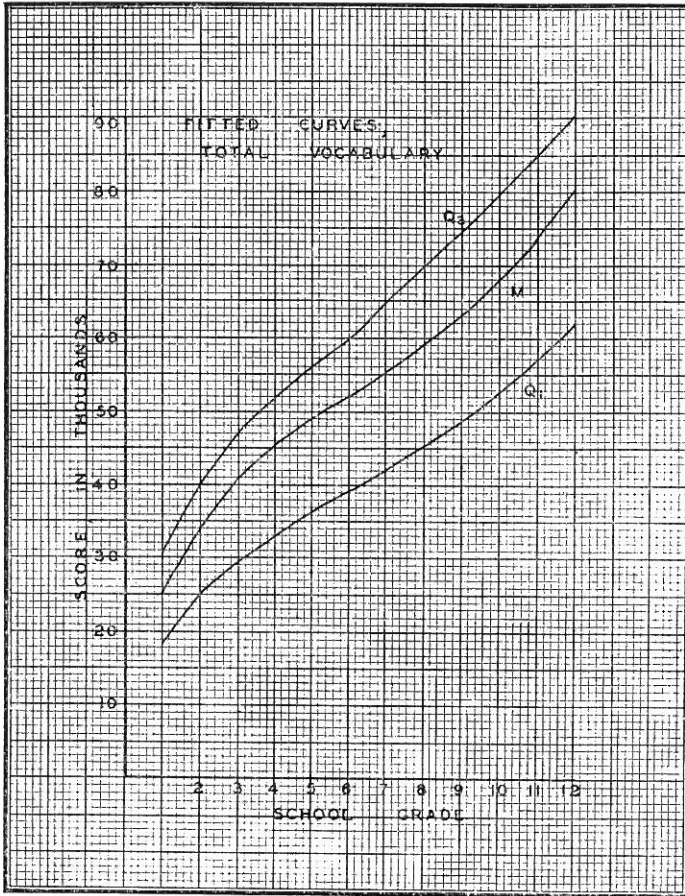


FIGURE 8

A child's score may be interpreted either in terms of the variability within his school grade or in terms of the averages of grades above or below his own. For example, the score of a fourth grader may be interpreted to mean that he ranks in the upper fourth of his class or that he has a vocabulary which is equal to the size of the average in the fifth grade.

Tentative age norms may be read from Figures 3 and 4.

A high degree of reliability in the testing is shown by the coefficients of reliability computed by correlating the scores on the even-numbered items with scores on the odd-numbered items in part one of the test.<sup>7</sup> Those uncorrected coefficients of reliability are shown in Table 4.

TABLE 4

| School grade | $r_{\frac{11}{22}}$ | PE  |
|--------------|---------------------|-----|
| 1            | .73                 | .05 |
| 2            | .75                 | .05 |
| 3            | .69                 | .05 |
| 4            | .75                 | .03 |
| 5            | .94                 | .01 |
| 6            | .91                 | .01 |
| 7            | .95                 | .01 |
| 8            | .92                 | .01 |
| 9            | .93                 | .01 |
| 10           | .87                 | .09 |
| 11           | .86                 | .02 |
| 12           | .76                 | .03 |

This study has shown the order of size of vocabulary for children in the elementary and high school. Many other studies of vocabulary could well follow this. Several questions concerning vocabulary growth for this same age range which remain unanswered are:

1. The factors which influence growth of vocabulary. These should be studied intensively. Suggested factors are:

- (a) Experience, trips, conversation
- (b) Wide reading
- (c) Systematic teaching of words—word drills, etc.
- (d) Analysis of words—teaching prefixes, suffixes, roots, etc.
- (e) Use situations—writing poetry, cross-word puzzles, cross-word lexicon, anagrams

<sup>7</sup>Reliability coefficients for grades seven to twelve inclusive also were computed by correlating odd vs. even items, counting as odd and even items the alternate words attempted rather than as the words were numbered in the test. Those uncorrected coefficients of reliability are: for

seventh grade,  $r_{\frac{11}{22}} = .90 \pm .02$ ; eighth grade,  $r_{\frac{11}{22}} = .92 \pm .01$ ; ninth grade,

$r_{\frac{11}{22}} = .91 \pm .01$ ; tenth grade,  $r_{\frac{11}{22}} = .87 \pm .02$ ; eleventh grade,  $r_{\frac{11}{22}} =$

$.87 \pm .02$ ; twelfth grade,  $r_{\frac{11}{22}} = .77 \pm .03$ .

It may be that these factors are of varying influence at different ages. There is evidence that direct experience is most important for children in the early grades whereas vicarious experience through reading is more valuable at the later ages.

2. There are differences between the child's various vocabularies such as those for recognition, reading, pronunciation, spelling, written use, and spoken use vocabularies. These differences need to be measured and the relationships between them studied.

3. There remains a problem of diction or selection of the *most appropriate* of several words beyond the rudimentary knowledge of any certain number of words. It should be determined how size of vocabulary relates to diction.

## V. SUMMARY

In summary, the Seashore-Eckerson *English Recognition Vocabulary Test*, designed to measure individual differences in total English vocabulary, was given to children from first grade through high school.

The test was given in two schools which had 12 grades in the school unit and in a third school which had only the first eight grades. The number of children tested at each grade level ranged from 40 at the second grade level to 114 in the ninth grade.

It was found that knowledge of words in the early grades was greatly affected by the methodology of testing and that improvement in vocabulary with age showed significant qualitative as well as quantitative changes.

Since we were interested in determining the total number of words which had any significant meaning for the child, a broad set of criteria of knowledge was adopted. Thus the child was first given an opportunity to define a word in his own terms or to illustrate its proper use in a sentence. If he was unable to meet these criteria he was then given an opportunity to demonstrate his ability to recognize the correct meaning on a four-choice multiple response test. At the earlier ages everything was read to the child to make sure that our results were not handicapped by inability to read, pronounce, or spell words. Standard, concise definitions were worked out for the alternative answer words of the multiple-choice questions and the students were instructed to ask for definitions of unfamiliar answer-words.

Grades one, two, and three were given individual oral tests exclusively; grade four was tested in small groups of four to five pupils with the examiner reading each item; while grades five and six were tested in larger groups, with the same method; grades seven through high school were tested in larger groups without the reading of the words by the examiner except as individual pupils asked for the pronunciation of certain words.

A fairly steady growth of vocabulary will be illustrated by the following figures. For grade one, the average number of basic words known was 16,900, with a range from 5,500 to 32,800. For grade twelve the average number of basic words known was 47,300, with a range from 28,200 to 73,200. For grade one the average number of words in the total vocabulary (basic plus derivative words) was

23,700, with a range from 6,000 to 48,800. For grade twelve the average number of words in the total vocabulary was 80,300, with a range from 36,700 to 136,500.

There was great overlapping of even the interquartile ranges of the different school grades. In basic vocabulary  $Q_1$  of the eighth grade exceeded  $Q_3$  of the first three grades. In terms of total vocabulary  $Q_1$  of the high school seniors was higher than  $Q_3$  of the seventh grade in the New Concord School and of the sixth grade in the Northbrook School, while  $Q_1$  of the eighth grade exceeded  $Q_3$  of only the first two grades.

An apparent dip in the curve for the growth of vocabulary at the fifth grade is believed to be due to the change in methodology brought about by testing in larger groups above the fourth grade, a procedure which has now been remedied. There is also some evidence for a genuine difference in ability at this level in one of the schools.

Although there were fairly large differences between the lower grades in different school systems, for the present purposes it seemed best to combine the results in a table of tentative norms, with quartile divisions, which may be interpreted either in terms of variability within the child's school grade or in terms of the averages of grades above or below his own. The same interpretations may be made in terms of chronological age.

## REFERENCES

1. GANSL, IRENE. Vocabulary: Its measurement and growth. *Arch. of Psychol.*, 1939, No. 236. Pp. 52.
2. LORGE, I. The English semantic count. *Teach. Coll. Rec.*, 1937, **39**, 65-77.
3. MCCARTHY, DOROTHEA. The Language Development of the Pre-School Child. Minneapolis: Univ. Minnesota Press, 1930.
4. SEASHORE, R. H., & ECKERSON, L. D. The measurement of individual differences in general English vocabularies. *J. Educ. Psychol.*, 1940, **31**, 14-38
5. SEASHORE, R. H., STOCKFORD, L. B. O., & SWARTZ, B. K. A correlational analysis of factors in speed of reading tests. *Sch. & Soc.*, 1937, **46**, 187-192.
6. SMITH, MADORAH E. An investigation of the development of the sentence and the extent of vocabulary in young children. *Univ. Iowa Stud. Child Wel.*, 1926, **3**, No. 5. Pp. 92.
7. TERMAN, L. M. The Measurement of Intelligence. New York: Houghton Mifflin, 1916.
8. THORNDIKE, E. L. A Teachers Word Book of the Twenty Thousand Words Found Most Frequently and Widely in General Reading for Children and Young People. New York: Bur. Publ., Teach. Coll., Columbia Univ., 1932.